Journal of Pioneering Medical Sciences
Received: May 05, 2025 | Accepted: June 08, 2025 | Published: September 05, 2025
Volume 14, Issue 08, Pages 200-212

DOI https://doi.org/10.47310/jpms2025140825

The Role of CYBB Gene Mutations in Chronic Granulomatous Disease: Exploring Immunogenicity and Immune Deficiencies

Wasan Abdulateef Majeed[1*], Israa Abdulqader Abdulwahab[2] and Shahad Abduljabbar Mohammed[3]
[1,3]Department of Biology, College of Education for Pure Science, University of Diyala, Iraq
[2]Ministry of Education, Resafa 1 Directorate of Education, Iraq

*Corresponding author: Wasan Abdulateef Majeed (e-mail: wassan.abdullateef@uodiyala.edu.iq).

©2025 the Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0

Abstract This study aims to characterise the structural impacts of missense mutations in the CYBB gene associated with Chronic Granulomatous Disease (CGD), by comparing them to wild-type CYBB and to assess their effects on protein stability and dynamics and thus on NADPH oxidase function. The goal is to elucidate the pathogenesis of CGD and inform personalised therapeutic strategies. The study used a computational approach combining data from the ClinVar database to identify and characterise CYBB missense variants. It included sequence conservation analysis, using multiple sequence alignments to identify conserved regions of the CYBB protein. DDMut, a deep-learning based stability assessment tool, was used to predict changes in Gibbs Free Energy (ΔΔG) upon mutation. Furthermore, molecular dynamics (MD) simulations were performed to investigate the stability and dynamics of CYBB and selected mutants. Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) data were analyzed to assess structural stability and flexibility, respectively. K-means clustering was used to group simulations. Analysis revealed numerous missense variants with varying pathogenicity classifications. Many variants classified as "likely benign" were predicted to destabilize the CYBB protein, while some "likely pathogenic" variants were predicted to increase stability. Mutations in conserved regions, including those classified as "likely benign," were found to be destabilizing. MD simulations revealed that some variants, despite maintaining overall structural similarity to the wild-type protein, induced substantial local dynamic alterations. Specific mutations within functional domains such as the transmembrane domain (e.g., S142P, R54G, A156D, I15V) and FAD and NAD-binding domains (e.g., P390L, W361R, G359R, E462A, C537F, K438R, Y476C, M465I) were predicted to be destabilizing. Some mutations outside the defined domains (C257R, F262C, I273V) were also found to be destabilizing. There are inconsistencies between ClinVar classifications and computational predictions regarding the impact of CYBB missense mutations on protein stability and dynamics. Many "likely benign" variants may have functional impacts and some "likely pathogenic" variants might not destabilize the protein. The study highlights the importance of using multiple computational approaches, including structural and dynamic analysis, to assess the impact of mutations in the CYBB gene. These findings contribute to a better understanding of the molecular mechanisms of CGD and will help in the development of more accurate diagnostic and therapeutic strategies. Experimental validation of these predictions is needed to further confirm these findings.

Key Words Chronic Granulomatous Disease (CGD), CYBB, Missense mutations, Protein stability and dynamics, NADPH oxidase

INTRODUCTION

Chronic Granulomatous Disease (CGD) is a group of inherited disorders that affect the immune system. Individuals with CGD have white blood cells (phagocytes) that do not function properly, making them more susceptible to recurrent and life-threatening bacterial and fungal infections. CGD is caused by mutations in any of the six genes that encode the subunits of the enzyme NADPH oxidase (CYBB, CYBA, CYBC1, NCF1, NCF2 and NCF4). Defects in NADPH oxidase 2 (Nox2), also known as cytochrome b(558) subunit beta or Cytochrome b-245 heavy chain (CYBB/NOX2) are the most common, accounting for approximately 70% of CGD cases. This protein is encoded by the *CYBB* gene located on the X chromosome. CGD follows an X-linked pattern of inheritance in those with a *CYBB* mutation. The estimated incidence of CGD is 1 in approximately 200,000 live births [1-3].

The cytochrome b-245 beta chain, a subunit of the NADPH oxidase complex, plays an essential role in the immune system. Within this complex, the cytochrome b-245

beta chain partners with an alpha chain (produced from the *CYBA* gene). Both alpha and beta chains are required for either to function and the NADPH oxidase complex requires both chains to be functional. The presence of foreign invaders stimulates phagocytes and triggers the assembly of NADPH oxidase. This enzyme participates in a chemical reaction that converts oxygen to a toxic molecule called superoxide. Superoxide is essential for the body's defence against bacteria and fungi [4].

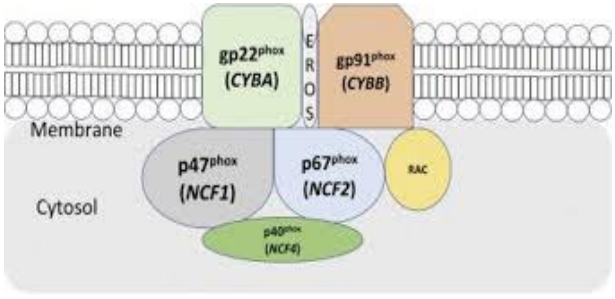
Individuals with CGD commonly experience infections in the lungs (pneumonia), lymph nodes (lymphadenitis), liver (abscess), bones (osteomyelitis) and skin (abscesses or cellulitis). Other manifestations include delayed growth, granulomas (especially in the genitourinary and gastrointestinal tracts), colitis, poor wound healing due to excessive granulation, growth restriction in childhood and chronic respiratory disease. Ophthalmic issues can include chorioretinal lesions and granulomata with pigment clumping. Oral manifestations may present as gingivitis, stomatitis, aphthous ulceration and gingival hypertrophy. Non-infectious skin problems can include photosensitivity, granulomatous lesions, vasculitis and excessive inflammation at wounds. Additionally, autoimmune disorders are also associated with CGD [1,5,6].

Over 650 different mutations in the CYBB gene are known to cause CGD. The majority of these mutations are single-nucleotide polymorphisms (SNPs), often resulting in a shortened, nonfunctional protein. These mutations include large and small deletions (11%), frameshift mutations (24%), nonsense mutations (23%), missense mutations (23%), splice-region mutations (17%) and regulatory-region mutations (2%). These mutations affect superoxide production, which in turn dictates the severity of CGD symptoms. Some mutations in *CYBB* result in no superoxide production, greatly reducing a person's ability to fight infection [8]. These mutations are associated with a more severe form of CGD, with individuals typically diagnosed earlier and experiencing higher mortality at a young age. In contrast, CGD, characterised by residual superoxide production, typically has a milder course and comes to clinical attention later in life [7,9].

Current treatment options for CGD caused by CYBB mutations include antibiotic and antifungal prophylaxis, interferon-gamma (IFN-γ) therapy, hematopoietic stem cell transplantation (HSCT) and gene therapy. Antibacterial prophylaxis with trimethoprim-sulfamethoxazole and antifungal prophylaxis with itraconazole help reduce infections but may lead to drug resistance and side effects. IFN-γ therapy boosts immune function, though its efficacy varies and it can cause flu-like symptoms. HSCT is the only known curative treatment, with over 90% survival in children under 14 but it carries risks such as graft-versus-host disease (GVHD) and transplant complications, with less favourable outcomes in older patients. Gene therapy is an emerging approach with curative potential but it remains experimental and poses risks like insertional mutagenesis and limited long-term data. Early diagnosis and treatment, particularly with HSCT, improve clinical outcomes [10].

While research explores therapies, the unclear pathogenicity of many CYBB mutations hinders progress. This ambiguity limits understanding of the genetic basis of CGD and the development of targeted therapies. While some mutations are clearly associated with disease, others lack definitive classification, limiting accurate diagnoses and prognoses. This highlights a critical research gap that needs attention to advance personalised medicine. A comprehensive characterisation of the functional consequences of these mutations is critically needed to elucidate their clinical significance, which may hinder the immune response. Such investigations will be instrumental in refining our understanding of the spectrum of NADPH oxidase dysfunction in CGD, particularly concerning variants currently classified as likely benign and ultimately improving diagnostic accuracy and personalised management strategies [6,11].

A comprehensive understanding of the structural consequences of all known ambiguous (likely benign and likely pathogenic) CYBB mutations associated with CGD remains elusive. This study addresses this gap by characterising the structural impacts of these CYBB mutations reported in the ClinVar database using advanced molecular modelling and dynamics techniques, coupled with deep learning-based structural stability assessment, in comparison to wild-type CYBB. We hypothesise that specific CYBB mutations will induce distinct structural perturbations within the protein, affecting its processing, folding and interaction with NADPH complex proteins. Through systematic analysis of the structural consequences of these mutations, this study aims to generate a valuable resource for elucidating the pathogenesis of CGD and ultimately contribute to the development of personalised therapeutic strategies. Furthermore, this work will computationally categorise previously undocumented mutations.



METHODS

Identification and Characterisation of Missense Variants in the CYBB Gene

To identify SNPs, specifically missense variants, within the CYBB gene, including their associated pathogenicity ("likely-benign" and "likely-pathogenic") classifications and corresponding amino acid substitutions, data were

retrieved from the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar) [17]. Furthermore, two SNPs classified as "pathogenic" and "benign" were randomly selected as controls to establish a comparative framework for classifying the true pathogenic or benign nature of the potentially likely benign variants. The wild-type amino acid sequence of human CYBB (P04839) was obtained from the UniProtKB database (https://www.uniprot.org/uniprotkb). *In silico* mutagenesis, substituting the wild-type residues with the identified variant amino acids, was performed using Maestro (Schrödinger Release 2023.To delineate the domain architecture of CYBB and assess the potential impact of the identified mutations on key functional regions, the Interpore database (https://www.ebi.ac.uk/interpro/), a resource encompassing protein families, domains and functional sites, was utilized [18,19].

Conservation Analysis of CYBB Across Multiple Species

To investigate the evolutionary trajectory of CYBB, a Multiple Sequence Alignment (MSA) was performed. CYBB protein sequences were retrieved from the UniProtKB database for the following species: Homo sapiens (P04839), Sus scrofa (P52649), Bos taurus (O46522), Rattus norvegicus (A6KU83), Mus musculus (Q61093), Cavia porcellus (H0UXC6), Gallus (A0A8V0X3X1), Danio rerio (Q7T2A7), Oryctolagus cuniculus (A0A5F9CCG5), Ovis aries (W5QC99), Canis lupus familiaris (A0A8C0SUZ1) and Pan troglodytes (H2R4R9).

The MSA was performed using Clustal Omega with default parameters, accessed through the European Bioinformatics Institute (EBI) web server (https://www.ebi.ac.uk/Tools/msa/clustalo/)[20]. The resulting alignment was then analysed for residue conservation using a custom Python script.

Residue conservation was assessed by calculating a conservation score for each position in the multiple sequence alignment. This score represents the frequency of the most prevalent amino acid at that position, expressed as a percentage. Based on these scores, residues were categorised as conserved (scores >80%), semi-conserved (50%≤ scores ≤80%) or non-conserved (scores <50%). To identify potential outliers within the conservation pattern, an Isolation Forest model, an unsupervised machine learning algorithm, was employed. Trained on the calculated conservation scores, this model identified anomalous residues. A conservation score of 1 indicates high conservation, reflecting strong amino acid preservation, while a score of 0 indicates no conservation. Protein sequence conservation reflects both ancestral and contemporary amino acid similarity, as well as the probability of evolutionary amino acid substitutions.

Predicting the Impact of CYBB Mutations on Protein Stability

To assess the impact of the selected CYBB mutations on protein stability, we employed DDMut, a fast and accurate Siamese neural network, to predict changes in Gibbs Free Energy (ΔΔG) upon mutation [21]. DDMut leverages graph-based representations of the localized 3D protein environment, combined with convolutional layers and transformer encoders, to effectively capture both short-range and long-range atomic interactions. This approach has demonstrated superior performance in predicting ΔΔG for SNPs, achieving high Pearson's correlations and outperforming existing methods, available as a web server (https://biosig.lab.uq.edu.au/ddmut). Following the assessment of stability changes induced by the CYBB mutations, variants were selected for further analysis based on their DDMut prediction scores. Mutations exhibiting a predicted change in ΔΔG of less than -0.3 kcal/mol were prioritized for in-depth investigation. Known pathogenic and benign variants were included as controls in the DDMut analysis to provide context for the predicted effects of the selected variants.

Comparative 3D Structural Analysis of Wild-type CYBB and Its Mutants

Following mutant structure generation, energy minimisation was performed using the CHARMM force field. This process, consisting of 100 steps, was executed using OpenMM, a toolkit designed for molecular dynamics simulations and protein model optimization [22]. This energy minimization step ensured protein structural stability and energetic favorability for subsequent analyses. PyMOL was used to visualise mutation-induced structural alterations by mapping protein domains onto their respective regions.

Molecular Dynamics Simulations of CYBB and Its Mutants

Molecular Dynamics (MD) simulations were performed using Maestro 12.0 (Schrödinger, LLC) to investigate the stability and dynamics of CYBB and the shortlisted unstable mutants based on ΔΔG. Protein structures were prepared, minimized and dehydrated using the Protein Preparation Wizard. Salt ions were added and the SPC force field was employed as the solvent model. Simulations were run for 50 ns at 300 K. Trajectory analysis, using the Interaction Diagram module, assessed conformational stability and dynamics, yielding Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) data, which reflect structural stability and flexibility, respectively. Using CYBB wild-type as a reference, MD simulation data were analyzed to investigate protein structure fluctuations by extracting alpha-carbon (Cα) atom coordinates from the trajectory files. RMSD and RMSF differences, quantifying the deviation of each simulation from the wild-type, were calculated. K-means clustering (k = 3) was performed on standardized RMSD and RMSF values to categorize simulations into "Highest," "Medium" and "Lowest" RMSF groups. Simulations exhibiting RMSD and RMSF values exceeding 2 Å were identified as displaying high fluctuations. All analyses were performed using a custom Python script.

RESULTS

Pathogenicity and Functional Analysis of CYBB Missense Mutations

Analysis of the ClinVar data revealed a series of missense variants within the CYBB gene, each with varying implications for pathogenicity. A total of 38 mutations were obtained from ClinVar reported for CGD. Notably, several mutations exhibited pathogenic, likely pathogenic, likely benign and unclassified classifications, suggesting a higher probability of varied contributing factors to disease development. However, we shortlisted all likely benign and likely pathogenic mutations (n = 36) along with one pathogenic (G359R) and one benign (G364R) variant as a control. These mutations, along with their classifications, are given in Table 1.

A range of protein changes have been observed and categorised based on their potential impact. A substantial portion of these variants are reported as likely benign, suggesting they have minimal to no effect on protein function or disease risk. These include changes such as I15V, V23I, R31W, S151N, A156D, K158R, R199K, A224T, I241L, Q246E, I273V, K318T, I325V, V358I, K438R, M465I, Y476C, N515S, N518K, S525C and R559Q indicating that benign variants are unlikely to disrupt ROS production. Conversely, a smaller subset of variants is classified as likely pathogenic, indicating that these changes are predicted to potentially disrupt the structure or function of the CYBB protein, possibly affecting NADPH oxidase activity and ROS production. These include N22K, R54G, H101Y, H115Y, S142P, H209R, C257R, F262C, W361R, T362I, P390L, G408A, P415R, E462A and C537F. Finally, one specific change, G359R, has been definitively classified as pathogenic, strongly suggesting a causative role in disease, while G364R is reported as a benign variant.

Concentration of Pathogenic Mutations in Conserved Regions of Cytochrome b-245 beta chain

Analysis of the MSA of CYBB with orthologous sequences from other species revealed a high degree of evolutionary conservation, with an overall sequence identity of 84.62%, as shown in Figure 1a-c. This indicates strong selective pressure maintaining the CYBB protein sequence across species. Despite this overall conservation, a significant proportion of mutations with varying functional impacts are located within these highly conserved regions. The majority of mutations in the conserved region are classified as either likely benign or likely pathogenic.

Table 1: Variants Identified in the CYBB Gene and Their Reported Classification

Protein change	Accession	Germline classification
G364R	VCV001575018	Benign
I15V	VCV001149798	Likely benign
V23I	VCV002170405	Likely benign
R31W	VCV002140468	Likely benign
S151N	VCV001086468	Likely benign
A156D	VCV001564687	Likely benign
K158R	VCV000968234	Likely benign
R199K	VCV001524475	Likely benign
A224T	VCV001982859	Likely benign
I241L	VCV001990428	Likely benign
Q246E	VCV001607220	Likely benign
I273V	VCV001093431	Likely benign
K318T	VCV001963950	Likely benign
I325V	VCV001564612	Likely benign
V358I	VCV000853858	Likely benign
K438R	VCV001081893	Likely benign
M465I	VCV001111272	Likely benign
Y476C	VCV000955030	Likely benign
N515S	VCV001589755	Likely benign
N518K	VCV001592208	Likely benign
S525C	VCV001558075	Likely benign
R559Q	VCV001142807	Likely benign
N22K	VCV000208503	Likely pathogenic
R54G	VCV002737183	Likely pathogenic
H101Y	VCV000010932	Likely pathogenic
H115Y	VCV000429474	Likely pathogenic
S142P	VCV001504854	Likely pathogenic
H209R	VCV000068399	Likely pathogenic
C257R	VCV001504925	Likely pathogenic
F262C	VCV001339531	Likely pathogenic
W361R	VCV002809842	Likely pathogenic
T362I	VCV000632782	Likely pathogenic
P390L	VCV000430219	Likely pathogenic
G408A	VCV001485565	Likely pathogenic
P415R	VCV000955588	Likely pathogenic
E462A	VCV001706598	Likely pathogenic
C537F	VCV002128256	Likely pathogenic
G359R	VCV003341071	Pathogenic

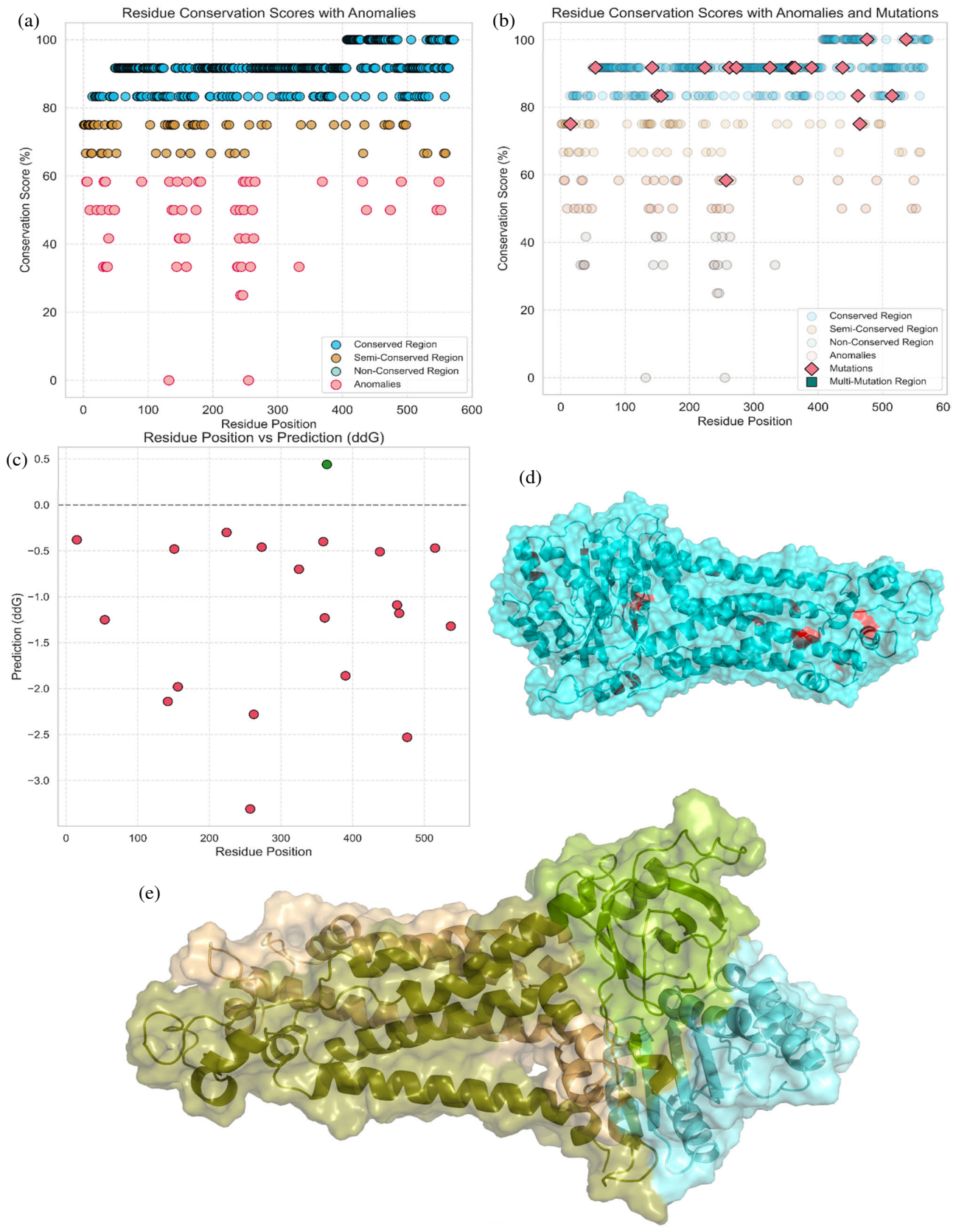


Figure 1(a-e): Conservation and location of CYBB gene mutations, (a-c) Conservation of the CYBB protein with orthologous sequences from other species, highlighting the high degree of evolutionary conservation, (d) Schematic representation of the CYBB protein showing the location of ClinVar-reported mutations, colored in red whereas cyan represents the protein and (e) Domain architecture of the CYBB protein, where green color represents the domain-less region

We observed most likely benign mutations in the conserved region, such as I151N, A156D, K158R, R199K, A224T, I273V, K318T, I325V, V358I, G364R (control/ benign), K438R, Y476C, N518K and S525C, which suggests a likely disruptive detrimental effect for the protein as opposed to their reported classification in ClinVar.

Similarly, likely pathogenic mutations like N22K, R54G, H101Y, H115Y, S142P, H209R, F262C, W361R, T362I, P390L, G408A, P415R, E462A and C537F are also observed in the conserved region of the protein. Interestingly, the control pathogenic mutation, G359R, is also observed in the conserved region suggesting to have a causative role in disease, likely by disrupting the protein's structure or function and affecting NADPH oxidase activity and ROS production. This indicates that these likely benign and pathogenic mutations should be analyzed further to determine the protein stability.

Most mutations in the semi-conserved region are classified as benign, including I15V, V23I, M465I and R559Q. There is only one likely pathogenic mutation in this region, C257R. All mutations in the non-conserved region, such as R31W, I241L and Q246E, are classified as benign, indicating that changes in this region are unlikely to disrupt the protein's function or ROS production.

Predicting the Impact of CYBB Mutations on Protein Stability

To assess the impact of the selected CYBB mutations on protein stability, we employed DDMut to predict changes in ΔΔG upon mutation. The DDMut analysis predicted destabilization of several CYBB mutants (ΔΔG<-0.3), likely impacting its function. A highly destabilising change is predicted for C257R (-3.31), where a Cysteine, often involved in disulfide bonds crucial for protein folding, is replaced by the bulkier and charged Arginine. This substitution could disrupt these bonds or introduce steric clashes.

Similarly, Y476C (-2.53) and F262C (-2.28), both involving changes to Cysteine, are predicted to be strongly destabilising, likely due to altered local interactions and disrupted hydrophobic packing. The S142P (-2.14) mutation, changing Serine to Proline, is also predicted to be destabilising due to Proline's rigid structure, which can disrupt secondary structures. Introducing a negative charge with the A156D (-1.98) mutation (Alanine to Aspartic acid) could lead to electrostatic repulsion or disrupt salt bridges. The P390L (-1.86) change (Proline to Leucine), while both hydrophobic, could still affect local protein structure. Another Cysteine change, C537F (-1.32), is predicted to be destabilising due to altered hydrophobic interactions. Replacing the large, positively charged Arginine with Glycine in R54G (-1.25) could disrupt electrostatic interactions. The W361R (-1.23) mutation (Tryptophan to Arginine) likely disrupts packing and electrostatic interactions. M465I (-1.18), E462A (-1.09), I325V (-0.7), K438R (-0.51), S151N (-0.48), N515S (-0.47), I273V (-0.46), G359R (-0.4) (the pathogenic control), I15V (-0.38) and A224T (-0.3) are also predicted to be destabilizing, with varying degrees of severity. This analysis suggests that DDMut correctly predicts the pathogenic variant to be unstable and the benign variant (G364R, 0.44) to be stable, while also indicating that the likely benign or pathogenic variants may be significantly unstable.

Mutations with smaller predicted impacts on stability (-0.3<ΔΔG<0) are considered neutral or slightly destabilizing. These changes, while potentially subtle, could still have functional consequences and warrant further investigation. Examples include, G408A and R199K, among others. Finally, some mutations are predicted to have a neutral or slightly stabilizing effect (ΔΔG≥0). K318T (0), H209R (0.08), R31W (0.1), S525C (0.11), I241L (0.21), Q246E (0.12), V23I (0.43) are predicted to have no destabilizing effect which are also classified as likely benign previously.

Furthermore, despite being classified as likely pathogenic on ClinVar, N22K (0.1), T362I (0.16), P415R (0.59), H101Y (1.55) and H115Y (2.12) are all predicted to slightly increase stability, therefore suggesting that these may not cause loss of function to the protein. Finally, the most destabilizing mutations, used for further analysis, are (in descending order of destabilization): C257R (Likely Pathogenic), Y476C/A156D (Likely Benign), F262C (Likely Pathogenic), S142P (Likely Pathogenic), P390L (Likely Pathogenic), C537F (Likely Pathogenic), R54G (Likely Pathogenic), W361R (Likely Pathogenic), M465I (Benign), E462A (Likely Pathogenic), I325V/K438R/ S151N/I273V (Likely Benign), G359R (Pathogenic), I15V/ A224T (Likely Benign), while G364R was further utilized a benign control, as confirmed by stability analysis. These shortlisted mutants are visualised in Figure 1c.

Mutational Analysis of the CYBB Protein within Functional Domains

Analysis of the CYBB revealed that it possesses a modular architecture characteristic of NADPH oxidases, featuring distinct domains responsible for its electron transfer function. A key feature is the Ferric reductase transmembrane component-like domain (residues 79-219). This region anchors the protein within a cellular membrane, a critical localization for its role in generating reactive oxygen species across that membrane.

The cytoplasmic portion of CYBB houses the catalytic machinery. It contains two overlapping FAD-binding domains: a more general FAD-binding 8 domain (residues 295-394) and a ferredoxin reductase-type FAD-binding domain (residues 287-397). These domains bind the FAD cofactor, essential for accepting electrons. Adjacent to the FAD-binding region is the Ferric reductase, NAD binding

domain (residues 401-550), which specifically binds NADPH. This domain serves as the entry point for electrons, with NADPH acting as the electron donor.

Finally, an overarching domain referred to as NOX_Duox_like_FAD_NADP, that encompasses residues 297-570. It is not a separate domain in terms of binding; it includes the aforementioned FAD-binding and NAD-binding domains. The overall structure positions the electron-accepting FAD near the membrane, facilitating electron transfer across the membrane, while the NADPH-binding domain remains cytoplasmic, providing the source of electrons. This arrangement is critical for function in the controlled generation of reactive oxygen species.

The mutations analyzed can be broadly categorized based on their location within the protein's structure and their impact. The first group involves mutations within or associated with, the Ferric reductase transmembrane component-like domain. This group includes S142P and R54G, both predicted to be unstable, indicating their potential to disrupt the transmembrane region's function. A156D and I15V/A224T within this region are also classified as unstable, suggesting a destabilizing impact on this transmembrane domain. S151N is also part of this domain, classified as unstable but part of a larger combination of mutations.

The second category encompasses mutations affecting the FAD-binding 8 domain (residues 295-394) and the Ferric reductase, NAD binding domain (residues 401-550). Within the FAD-binding domain, P390L, W361R and G359R are all unstable, highlighting the sensitivity of this region to amino acid changes. G364R, also in this domain, serves as a stable control, which is interesting as it indicates that not all residues are prone to destabilizing effect on the protein structure despite being in the conserved or domain region. I325V falls into the larger combination classification. The NAD-binding domain shows a similar pattern where E462A, C537F and K438R Y476C and M465I are predicted to be destabilizing mutants.

Finally, a third category includes mutations located outside of the specifically defined domains. This group contains C257R and F262C, both classified as unstable, suggesting that regions outside the well-defined domains are also critical for protein stability. I273V resides outside the defined region and is classified as unstable. This distribution indicates that destabilizing mutations are not limited to known functional domains but are spread across various locations. The mutations are depicted in Figure 1D whereas the protein domains are depicted in Figure 1e.

Analysis of Cytochrome b-245 beta chain Variant Stability and Dynamics

The analysis of the MD simulations revealed several cytochrome b-245 beta chain variants with higher RMSF values compared to the wild-type CYBB protein. The cluster with lowest RMSF (<0.6) revealed that C257R, G364R, K438R, N515S, S151N, W361R and Y476C fall into this category. These mutants exhibit the least fluctuation in their residues compared to the other groups. Crucially, the results indicate that these have the lowest deviation from the wild type. This strongly suggests that these mutations have minimal impact on the overall flexibility and dynamics of the CYBB protein, closely resembling the wild-type's behavior.

The cluster with medium RMSF (0.6<RMSF<0.82) revealed that A224T, C537F, E462A and M465I are in this group. These mutants show a moderate level of fluctuation. Their RMSF values are higher than the "Lowest" group, indicating increased flexibility compared to the wild-type and the lowest RMSF cluster. This suggests that the mutations in these mutants might be causing some local structural changes or influencing the dynamics of nearby residues, leading to greater movement. These mutations could be in regions that are intrinsically more flexible or they might be disrupting some interactions but not to a drastic extent, as depicted in Figure 2b and 2d.

Finally, the cluster with highest RMSF (0.6<RMSF< 0.82) revealed that A156D, F262C, G359R, I15V, I273V, I325V, P390L, R54G and S142P, belong to the highest group. This increased fluctuation could have functional consequences. For mutants in the highest group, they are on the higher end. Although they have low RMSF scores, they might still indicate a level of deviation from wild-type.

Based on the K-means clustering of the RMSD values, we observed three different clusters. The mutants G364R, I325V, N515S, P390L and S151N belong to the lowest RMSD cluster, exhibiting RMSD values ranging from 0.49 Å to 0.69 Å. These mutations exhibit minimal structural deviations from the reference structure, suggesting that their impact on the overall protein conformation is negligible. The low RMSD values indicate that the backbone conformation and most side-chain orientations in these mutants closely resemble those of the reference. These mutations likely occur in regions that are either not critical for maintaining the protein’s fold or involve conservative substitutions, leading to minimal disruptions in the protein’s overall structure, as depicted in Figure 2a and 2c.

The medium RMSD cluster, with RMSD values ranging from 0.77 Å to 1.05 Å, includes the mutants A156D, I273V, K438R, R54G, S142P, W361R and Y476C. These mutants display a moderate degree of structural deviation, with RMSD values higher than those in the "Lowest" cluster but still within a relatively low range. This suggests that while the overall protein fold is largely preserved, there are more noticeable structural differences compared to the "Lowest" group. These deviations may manifest as localized conformational changes, such as shifts in loop regions or alterations in side-chain packing, indicating that the mutations have a moderate impact on the protein’s structure.

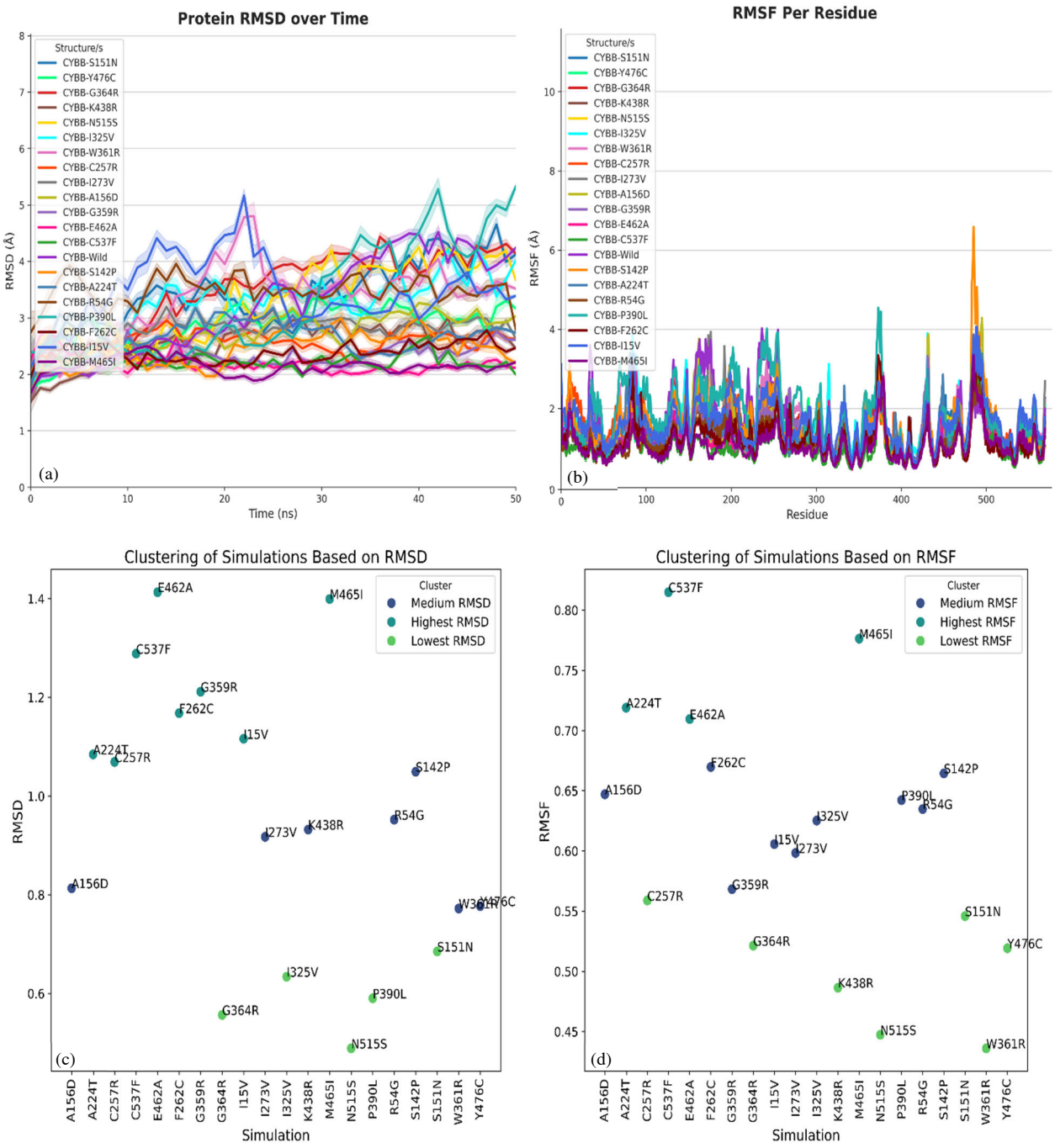


Figure 2(a-d): Structural dynamics and stability of CYBB variants, (a) RMSD of the Cα atoms of the CYBB variants throughout the 50 ns MD simulations, relative to the starting structure, (b) RMSF of the Cα atoms of the CYBB variants throughout the 50 ns MD simulations, relative to the wild-type CYBB protein, (c) K-means clustering of the RMSD values, showing the grouping of variants based on their structural deviation from the wild type and (d) K-means clustering of the RMSF values, showing the grouping of variants based on their flexibility compared to the wild type

Finally, the highest RMSD cluster, with RMSD values ranging from 1.07 Å to 1.41 Å, includes the mutants A224T, C257R, C537F, E462A, F262C, G359R, I15V and M465I. These mutants exhibit the largest structural deviations from the reference structure, although their RMSD values remain below the 2 Å threshold. The noticeable structural changes in this group suggest that the mutations have a more substantial impact on the protein's overall conformation. These deviations may involve significant movements in loop regions, domain shifts or alterations in the packing of secondary structural elements. While the overall fold likely remains similar to the reference, the structural changes in these mutants are more pronounced compared to the other clusters. Finally, when compared to the original germline classification, we propose a new MD-simulations-based classification of the CYBB variants, as given in Table 2.

Table 2: Classifying CYBB Variants

Mutant	Original Classification (ClinVar)	Conservation Classification	DDMut Classification	Domain Location	MD Simulation Classification (RMSF/RMSD)
G364R	Benign	Conserved	Stable	FAD-binding 8	Stable (Lowest RMSF & RMSD)
I15V	Likely Benign	Semi-Conserved	Destabilizing	Ferric reductase transmembrane component-like	Unstable (Highest RMSD)
V23I	Likely Benign	Semi-Conserved	Neutral	None	Skipped
R31W	Likely Benign	Non-Conserved	Neutral	None	Skipped
S151N	Likely Benign	Conserved	Destabilizing	Ferric reductase transmembrane component-like	Stable (Lowest RMSF & RMSD)
A156D	Likely Benign	Conserved	Destabilizing	Ferric reductase transmembrane component-like	Unstable (Medium RMSD, Highest RMSF)
K158R	Likely Benign	Conserved	Neutral	None	Skipped
R199K	Likely Benign	Conserved	Neutral	None	Skipped
A224T	Likely Benign	Conserved	Destabilizing	Ferric reductase transmembrane component-like	Unstable (Highest RMSD, Medium RMSF)
I241L	Likely Benign	Non-Conserved	Neutral	None	Skipped
Q246E	Likely Benign	Non-Conserved	Neutral	None	Skipped
I273V	Likely Benign	Conserved	Destabilizing	None	Unstable (Medium RMSD)
K318T	Likely Benign	Conserved	Neutral	None	Skipped
I325V	Likely Benign	Conserved	Destabilizing	FAD-binding 8	Stable (Lowest RMSD),Unstable (Highest RMSF)
V358I	Likely Benign	Conserved	Neutral	None	Skipped
K438R	Likely Benign	Conserved	Destabilizing	Ferric reductase, NAD binding	Stable (Lowest RMSF), Unstable (Medium RMSD)
M465I	Likely Benign	Semi-Conserved	Destabilizing	Ferric reductase, NAD binding	Unstable (Highest RMSD, Medium RMSF)
Y476C	Likely Benign	Conserved	Destabilizing	Ferric reductase, NAD binding	Stable (Lowest RMSF),Unstable (Medium RMSD)
N515S	Likely Benign	Conserved	Destabilizing	Ferric reductase, NAD binding	Stable (Lowest RMSF & RMSD)
N518K	Likely Benign	Conserved	Neutral	None	Skipped
S525C	Likely Benign	Conserved	Neutral	None	Skipped
R559Q	Likely Benign	Semi-Conserved	Neutral	None	Skipped
N22K	Likely Pathogenic	Conserved	Stable	None	Skipped
R54G	Likely Pathogenic	Conserved	Destabilizing	Ferric reductase transmembrane component-like	Unstable (Medium RMSD, Highest RMSF)
H101Y	Likely Pathogenic	Conserved	Stable	None	Skipped
H115Y	Likely Pathogenic	Conserved	Stable	None	Skipped
S142P	Likely Pathogenic	Conserved	Destabilizing	Ferric reductase transmembrane component-like	Unstable (Medium RMSD, Highest RMSF)
H209R	Likely Pathogenic	Conserved	Neutral	None	Skipped
C257R	Likely Pathogenic	Semi-Conserved	Destabilizing	None	Unstable (Highest RMSD, Lowest RMSF)
F262C	Likely Pathogenic	Conserved	Destabilizing	None	Unstable (Highest RMSD, Highest RMSF)
W361R	Likely Pathogenic	Conserved	Destabilizing	FAD-binding 8	Stable (Lowest RMSF),Unstable (Medium RMSD)
T362I	Likely Pathogenic	Conserved	Stable	FAD-binding 8	Skipped
P390L	Likely Pathogenic	Conserved	Destabilizing	FAD-binding 8	Stable (Lowest RMSD), Unstable (Highest RMSF)
G408A	Likely Pathogenic	Conserved	Neutral	Ferric reductase, NAD binding	Skipped
P415R	Likely Pathogenic	Conserved	Stable	Ferric reductase, NAD binding	Skipped
E462A	Likely Pathogenic	Conserved	Destabilizing	Ferric reductase, NAD binding	Unstable (Highest RMSD, Medium RMSF)
C537F	Likely Pathogenic	Conserved	Destabilizing	Ferric reductase, NAD binding	Unstable (Highest RMSD, Medium RMSF)
G359R	Pathogenic	Conserved	Destabilizing	FAD-binding 8	Unstable (Highest RMSF)

DISCUSSION

This study is the first to explore the effects of missense variants (classified as likely-benign and likely-pathogenic) in the CYBB gene, which encodes the cytochrome b-245 beta chain, a crucial subunit of the NADPH oxidase complex, on protein stability and dynamics. Using an integrated computational approach that combines ClinVar classifications, sequence conservation analysis, DDMut stability predictions and MD simulations, we uncovered notable discrepancies, offering a deeper insight into the potential roles of these variants in CGD. A key finding was the frequent inconsistency between ClinVar classifications

and the results from our computational analyses. While many variants were categorized as "likely benign," our predictions suggested they might destabilize the protein, possibly affecting CYBB function. In contrast, some "likely pathogenic" variants were predicted to enhance protein stability, prompting further investigation into their underlying pathogenic mechanisms. Significantly, MD simulations revealed that variants, despite maintaining overall structural similarity to the wild-type protein, could induce substantial local dynamic alterations, suggesting the limitations of relying solely on static structural assessments.

The discrepancies observed between ClinVar classifications and our computational predictions highlight the complexities of genotype-phenotype correlations. Several factors contribute to these inconsistencies. ClinVar, while an invaluable resource, is based on clinical data submissions that can vary in quality, completeness and supporting evidence. Its primary focus is on the clinical significance of variants, specifically whether they are associated with disease, which does not always correlate directly with their molecular impact on protein stability or dynamics. A variant may be clinically benign if the residual protein function remains sufficient, even in the presence of some destabilization. Additionally, in vivo compensatory mechanisms, such as interactions with other cellular proteins or pathways, could mitigate the effects of a destabilizing mutation, resulting in a less severe or even absent clinical phenotype. ClinVar classifications may also be influenced by factors outside the coding sequence analyzed in our study, such as regulatory region variations that affect gene expression. Finally, incomplete penetrance and variable expressivity, which are common in genetic disorders, can further complicate the determination of pathogenicity based solely on clinical observations. The CYBB synonymous variant c.483G>A exemplifies how ClinVar can be misleading. Initially a VUS, this seemingly silent change was later shown to affect splicing, demonstrating that even seemingly benign variants can have significant functional consequences. This case highlights the danger of relying solely on ClinVar, especially for synonymous variants and underscores the need for functional studies to accurately interpret variant pathogenicity [14,23].

For instance, Y476C, located in the NAD-binding domain and classified as "likely benign" in ClinVar, was predicted to be highly destabilizing by DDMut (-2.53 ΔΔG) and exhibited increased residue fluctuation (lowest RMSF, medium RMSD) in MD simulations. This suggests a potential disruption of NADPH binding or electron transfer. The discrepancy could arise from limited clinical data associated with this specific variant, potential compensatory mechanisms in vivo that maintain sufficient NADPH oxidase activity or the mutation's impact being primarily on a region not directly involved in catalysis but crucial for overall structural integrity. Similarly, A156D, located in the transmembrane domain, also classified as "likely benign," was predicted to be destabilizing (-1.98 ΔΔG) and showed increased flexibility (highest RMSF, medium RMSD). This could compromise membrane anchoring, essential for CYBB's function.

The NADPH-binding domain, located within the C-terminal cytoplasmic region of the gp91phox protein (encoded by the CYBB gene), is essential for the catalytic activity of the NADPH oxidase complex. This domain, crucial for binding the NADPH substrate, facilitates electron transfer and the subsequent generation of ROS, critical for microbial killing. Mutations within this domain, including missense mutations, small deletions and insertions, can disrupt protein structure and function, often resulting in decreased or abolished NADPH oxidase activity. Missense mutations can alter critical amino acid residues within the binding site, while deletions and insertions can lead to frameshifts and truncated proteins lacking the essential binding site. Specific examples, such as the Leu505Arg mutation (impacting adenine binding), the Gly412Glu mutation (distorting the Rossman fold) and mutations at Pro415 (disrupting NADPH binding), demonstrate the diverse mechanisms by which these mutations can impair NADPH binding and ROS production. Furthermore, mutations affecting residues near the NADPH binding site, such as Thr503Ile or larger deletions encompassing the region, can also disrupt function. These mutations can manifest clinically as various forms of CGD, including X91+ CGD, often presenting with severe infections due to impaired phagocyte function [23].

Conversely, N22K and H115Y, both classified as "likely pathogenic," were predicted to increase stability by DDMut (0.1 and 2.12 ΔΔG, respectively). This suggests that their pathogenic mechanism may not involve destabilization of the protein's overall fold. Instead, these mutations might disrupt specific protein-protein interactions, alter the protein's subcellular localization or affect its regulation, mechanisms not directly captured by our stability and dynamics analyses [24].

Disagreements were also observed between conservation analysis and stability/dynamics predictions. While mutations in highly conserved regions are often expected to be deleterious, this isn't always the case. Conservation, while high overall (84.62% sequence identity), isn't absolute. Some positions can tolerate conservative substitutions. For example, I325V, located in the highly conserved FAD-binding domain, was classified as likely benign and found in a conserved area. While DDMut predicted destabilization, the impact of I325V it also had among the lowest RMSD score, suggesting that this position could potentially handle conservative substitution, however it might affect local residue which are important for function. Furthermore, DDMut's focus on the immediate vicinity of the mutation might miss long-range effects, while MD simulations, although limited in timescale, can reveal more global changes in flexibility [16,25,26].

The comparison between DDMut and MD results also revealed interesting nuances. A mutation predicted to be destabilizing by DDMut might show relatively low RMSD in MD simulations, suggesting that the overall protein fold is maintained. However, RMSF analysis could reveal increased flexibility in specific regions, indicating a localized structural perturbation that could still affect function. This highlights the importance of considering both overall stability (RMSD) and local dynamics (RMSF).

The location of the mutations within specific protein domains provided further insights. Destabilizing mutations within the transmembrane domain (e.g., S142P, R54G, A156D, I15V) could impair the protein's anchoring within the membrane, crucial for its function in generating reactive oxygen species across the membrane. Mutations within the FAD- and NAD-binding domains (e.g., P390L, W361R, G359R, E462A, C537F, K438R, Y476C, M465I) could directly interfere with cofactor binding and electron transfer, disrupting the core catalytic activity of the enzyme. Interestingly, mutations outside these well-defined domains (e.g., C257R, F262C, I273V) were also predicted to be destabilizing, emphasizing that regions beyond known active sites or binding pockets are critical for maintaining overall protein stability and function.

Mutations in the six transmembrane domains, essential for anchoring gp91phox to the cell membrane and interacting with other NADPH oxidase components, can impair electron transfer and ROS production. Specific examples include the c.55C>G (p.Leu19Val) mutation, potentially disrupting the protein's helical structure and mutations like p.Arg54Ser and p.Ala57Glu, affecting heme redox potential and electron exit. Similarly, mutations within the FAD-binding domain, crucial for electron transfer from NADPH to oxygen, can disrupt FAD binding. Examples like Val327Asp and mutations at P390L, W361R and G359R illustrate how these changes can impair FAD binding and consequently ROS production. Mutations affecting the NADPH-binding domain, essential for binding the primary electron donor, also disrupt electron transfer and ROS generation. The Leu505Arg, Gly412Glu and Pro415His mutations exemplify how these changes can hinder NADPH binding and abolish or severely reduce NADPH oxidase activity. Furthermore, mutations like Thr503Ile and deletions within this region can also disrupt the binding site. The clinical consequences of CYBB mutations vary, leading to different CGD phenotypes (X910, X91- and X91+), characterized by varying levels of gp91phox expression and NADPH oxidase activity.

Destabilizing mutations, especially those located in conserved regions and functional domains of the CYBB gene, are likely to impair its function by disrupting electron transfer, reducing ROS production or affecting membrane association. These effects align with CGD's pathophysiology, which is characterized by defective NADPH oxidase activity and an impaired ability of phagocytes to kill pathogens. Variants with conflicting evidence complicate the interpretation of their roles. Although classified as "likely benign" in ClinVar, many of these variants showed destabilizing effects in our computational analyses, suggesting that they might have subtle, functionally relevant impacts on CYBB. These variants could contribute to variations in disease severity or predispose individuals to other health issues, warranting further research to clarify their roles.

Mutations affecting the FAD-binding and NADPH-binding domains, such as Leu505Arg, Gly412Glu, Pro415His and Val327Asp, directly impair the binding of essential cofactors (NADPH and FAD) and thus hinder electron flow. Similarly, mutations like the one at His209, involved in heme coordination, also disrupt electron transfer. These disruptions directly lead to reduced ROS production, the key effector mechanism for pathogen killing in phagocytes. Mutations such as p.Thr208Arg (impacting heme incorporation) and c.55C>G (p.Leu19Val) (potentially damaging helical structure) further contribute to reduced ROS production. Furthermore, mutations within the transmembrane domains, like p.Arg54Ser and p.Ala57Glu, can impair electron transport across the membrane. These transmembrane domains are also crucial for anchoring gp91phox to the phagocyte membrane and mutations like L19V can disrupt protein structure and membrane association, hindering proper NADPH oxidase complex assembly and function. Computational analyses often reveal the destabilising effects of these mutations, highlighting their impact on protein folding, stability and interactions. For example, p. Thr208Arg and p. Val327Asp are predicted to have significant structural consequences [23,34].

The few variants predicted to enhance stability raise important questions. While increased rigidity may seem beneficial, it could hinder the protein's ability to undergo essential conformational changes for optimal function. This could interfere with protein-protein interactions or alter the protein's susceptibility to regulatory modifications, which may ultimately have detrimental effects on CYBB function.

Our findings could help explain the variability in disease severity observed among CGD patients. The different mutations, with varying effects on CYBB stability and dynamics, may result in different levels of NADPH oxidase dysfunction, which could contribute to the diverse clinical presentations seen in CGD.

This study has several limitations. The DDMut and MD simulations, while powerful, are based on approximations of interatomic forces and may not fully capture the complexities of protein behavior *in vivo*. The limited timescale of MD simulations also prevents observation of slow conformational changes. Additionally, experimental validation is crucial to confirm our predictions and our study did not account for the cellular context or potential interactions with other proteins and genetic modifiers. Future research should prioritize the experimental validation

of our computational predictions, including in vitro assays to measure CYBB activity and ROS production. Structural studies and site-directed mutagenesis could provide further insights into the effects of mutations. Studies in cell culture and animal models are necessary to explore compensatory mechanisms and improving computational methods for variant impact prediction will be valuable. Large-scale clinical studies correlating genotype with phenotype will also help clarify the relationship between specific CYBB mutations and disease severity.

This study indicates the importance of integrating computational approaches to predict the functional impact of missense variants in CYBB. Our analysis, which combines sequence conservation, stability predictions and molecular dynamics simulations, identified variants likely to disrupt protein function and contribute to CGD pathogenesis, even if classified as "likely benign." The discrepancies between our predictions and existing clinical observations highlight the complexity of genotype-phenotype relationships and the need for further experimental and clinical investigations to fully understand CGD's molecular mechanisms and improve diagnostic and therapeutic strategies.

CONCLUSIONS

This study reveals discrepancies between ClinVar classifications and computational predictions of CYBB variant effects on protein stability, highlighting the limitations of relying solely on clinical data. Integrating computational approaches, including molecular dynamics, is crucial for assessing CYBB mutation impact, revealing that even seemingly benign variants can induce significant dynamic changes. Many "likely benign" ClinVar variants were predicted to be destabilizing, while some "likely pathogenic" variants were predicted to be stabilizing, necessitating re-evaluation of classifications. Destabilizing mutations were identified within and outside functional domains, emphasizing the importance of overall CYBB stability. These findings enhance our understanding of CGD molecular mechanisms and have implications for improved diagnostics and therapies. While computational predictions are valuable, experimental validation remains essential. This work also helps explain the variability in CGD severity, as different mutations exert varying effects on CYBB stability. In conclusion, a comprehensive approach integrating computational, experimental and clinical data is essential for accurate CYBB variant assessment and a deeper understanding of CGD.

REFERENCES

[1] Mollin, M. *et al.* "Clinical, Functional and Genetic Characterization of 16 Patients Suffering from Chronic Granulomatous Disease Variants - Identification of 11 Novel Mutations in CYBB." *Clinical & Experimental Immunology*, vol. 203, 2021, pp. 247-266.

[2] Di Matteo, G. *et al.* "Molecular Characterization of a Large Cohort of Patients with Chronic Granulomatous Disease and Identification of Novel CYBB Mutations: An Italian Multicenter Study." *Molecular Immunology*, vol. 46, 2009, pp. 1935-1941.

[3] Rae, J. *et al.* "X-Linked Chronic Granulomatous Disease: Mutations in the CYBB Gene Encoding the gp91-phox Component of Respiratory-Burst Oxidase." *American Journal of Human Genetics*, vol. 62, 1998, pp. 1320-1331.

[4] Jirapongsananuruk, O. *et al.* "CYBB Mutation Analysis in X-Linked Chronic Granulomatous Disease." *Clinical Immunology*, vol. 104, 2002, pp. 73-76.

[5] Arnold, D.E. and J.R. Heimall. "A Review of Chronic Granulomatous Disease." *Advances in Therapy*, vol. 34, 2017, pp. 2543-2557.

[6] Martel, C. *et al.* "Clinical, Functional and Genetic Analysis of Twenty-Four Patients with Chronic Granulomatous Disease - Identification of Eight Novel Mutations in CYBB and NCF2 Genes." *Journal of Clinical Immunology*, vol. 32, 2012, pp. 942-958.

[7] Segal, A.W. *et al.* "Absence of Cytochrome b-245 in Chronic Granulomatous Disease." *New England Journal of Medicine*, vol. 308, 1983, pp. 245-251.

[8] Azarova, I.E. *et al.* "Polymorphisms of the Gene Encoding Cytochrome b-245 Beta Chain of NADPH Oxidase: Relationship with Redox Homeostasis Markers and Risk of Type 2 Diabetes Mellitus." *Russian Journal of Genetics*, vol. 56, 2020, pp. 856-862.

[9] Teahan, C. *et al.* "The X-Linked Chronic Granulomatous Disease Gene Codes for the β-Chain of Cytochrome b–245." *Nature*, vol. 327, 1987, pp. 720-721.

[10] Staudacher, O. and H. von Bernuth. "Clinical Presentation, Diagnosis and Treatment of Chronic Granulomatous Disease." *Frontiers in Pediatrics*, vol. 12, 2024.

[11] Gérard, B. *et al.* "Characterization of 11 Novel Mutations in the X-Linked Chronic Granulomatous Disease (CYBB Gene)." *Human Mutation*, vol. 18, 2001, pp. 163-163.

[12] Stasia, M.J. *et al.* "Characterization of Six Novel Mutations in the CYBB Gene Leading to Different Sub-Types of X-Linked Chronic Granulomatous Disease." *Human Genetics*, vol. 116, 2005, pp. 72-82.

[13] Reis, B.C.S. *et al.* "Chronic Granulomatous Disease and Myelodysplastic Syndrome in a Patient with a Novel Mutation in CYBB." *Genes*, vol. 12, 2021, p. 1476.

[14] Roos, D. *et al.* "Hematologically Important Mutations: The Autosomal Forms of Chronic Granulomatous Disease (Third Update)." *Blood Cells, Molecules and Diseases*, vol. 92, 2021, p. 102596.

[15] Gul, I. *et al.* "Novel Mutations in CYBB Gene Cause X-Linked Chronic Granulomatous Disease in Pakistani Patients." *Italian Journal of Pediatrics*, vol. 49, 2023, p. 95.

[16] Wu, E.Y. *et al.* "Expanding the Clinical Phenotype of Chronic Granulomatous Disease: A Female Patient with a De Novo Mutation in CYBB." *Journal of Clinical Immunology*, vol. 41, 2021, pp. 224-226.

[17] "ClinVar: Public Archive of Interpretations of Clinically Relevant Variants." *Nucleic Acids Research*, 2016. https://academic.oup.com/nar/article/44/D1/D862/2502702.

[18] "InterPro in 2022." *Nucleic Acids Research*, 2022. https://academic.oup.com/nar/article/51/D1/D418/6814474.

[19] "UniProt: A Hub for Protein Information." *Nucleic Acids Research*, 2015. https://academic.oup.com/nar/article/43/D1/D204/2439939.

[20] Sievers, F. and D.G. Higgins. "Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences." *Multiple Sequence Alignment Methods*, edited by D.J. Russell, Humana Press, 2014, pp. 105-116. doi:10.1007/978-1-62703-646-7_6.

[21] "DDMut: Predicting Effects of Mutations on Protein Stability Using Deep Learning." *Nucleic Acids Research*, 2023. https://academic.oup.com/nar/article/51/W1/W122/7191416.

[22] "OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials." *Journal of Physical Chemistry B*, 2023. https://pubs.acs.org/doi/abs/10.1021/acs.jpcb.3c06662.

[23] Buratti, E. *et al.* "Aberrant 5' Splice Sites in Human Disease Genes: Mutation Pattern, Nucleotide Structure and Comparison of Computational Tools That Predict Their Utilization." *Nucleic Acids Research*, vol. 35, 2007, pp. 4250-4263.

[24] Byrum, S. *et al.* "Analysis of Stable and Transient Protein-Protein Interactions." *Methods in Molecular Biology*, vol. 833, 2012, pp. 143-152.

[25] Zhang, J. *et al.* "Identification of a Novel Mutation in CYBB Gene in a Chinese Neonate with X-Linked Chronic Granulomatous Disease: A Case Report." *Medicine (Baltimore)*, vol. 101, 2022, e28875.

[26] Yu, L. *et al.* "De Novo Somatic Mosaicism of CYBB Caused by Intronic LINE-1 Element Insertion Resulting in Chronic Granulomatous Disease." *Journal of Clinical Immunology*, vol. 43, 2023, pp. 88-100.

[27] Sevim-Wunderlich, S. *et al.* "A Mouse Model of X-Linked Chronic Granulomatous Disease for the Development of CRISPR/Cas9 Gene Therapy." *Genes*, vol. 15, 2024, p. 706.

[28] Zerbe, C.S. and S.M. Holland. "Functional Neutrophil Disorders: Chronic Granulomatous Disease and Beyond." *Immunological Reviews*, vol. 322, 2024, pp. 71-80.

[29] Chen, X. *et al.* "Preimplantation Genetic Testing for X-Linked Chronic Granulomatous Disease Induced by a CYBB Gene Variant: A Case Report." *Medicine (Baltimore)*, vol. 103, 2024, e37198.

[30] Wilson, B.E. *et al.* "Novel Intronic Variant in CYBB Causing X-Linked Chronic Granulomatous Disease: Case Report." *Medical Reports*, vol. 5, 2024, p. 100060.

[31] "A Novel Missense Mutation in CYBB Gene in Chronic Granulomatous Disease." *Journal of Allergy and Clinical Immunology*, 2006. https://www.jacionline.org/article/S0091-6749(06)03489-0/fulltext.

[32] "Variable Presentation of the CYBB Mutation in One Family, Approach to Management and a Review of the Literature." *Case Reports in Medicine*, 2020. https://onlinelibrary.wiley.com/doi/10.1155/2020/2546190.

[33] El Kares, R. *et al.* "Genetic and Mutational Heterogeneity of Autosomal Recessive Chronic Granulomatous Disease in Tunisia." *Journal of Human Genetics*, vol. 51, 2006, pp. 887-895.

[34] de Albuquerque, J.A.T. *et al.* "A C126R de Novo Mutation in CYBB Leads to X-Linked Chronic Granulomatous Disease with Recurrent Pneumonia and BCGitis." *Frontiers in Pediatrics*, vol. 6, 2018.